



DETECTING DNA SEQUENCE VARIATIONS

INTRODUCTION

Detection of variation in DNA sequences forms the basis of many applications in modern genetic analysis: it is used in linkage analysis to track disease genes in human pedigrees or economically important traits in animal and plant breeding programmes; it forms the basis of

10 fingerprinting methods used in forensic and paternity testing [Krawczak and Schmidtke, 1994]; it is used to discover mutations in biologically and clinically important genes [Cooper and Krawczak, 1989]. The importance of DNA polymorphism is underlined by the large number of methods that have been developed to detect and measure it [Cotton, 1993]. Most of

15 these methods depend on one of two analytical procedures, gel electrophoresis or molecular reassociation, to detect sequence variation. Each of these powerful procedures has its drawbacks. Gel electrophoresis has very high resolving power, and is especially useful for the detection of variation in the mini- and microsatellite markers that are used in linkage

20 analysis and fingerprinting; it is also the method used to analyse the variation found in the triplet repeats that cause a number of mutations now known to be the cause of around ten genetic disorders in humans [Willems, 1994]. Despite its great success and widespread use, gel electrophoresis has proved difficult to automate: even the systems which

25 automate data collection require manual gel preparation; and as samples are loaded by hand, it is easy to confuse samples. The continuous reading electrophoresis machines are expensive, and manual analysis is technically demanding, so that its use is confined to specialised laboratories which have a high throughput. Furthermore, difficulties in

30 measuring fragment size preclude rigorous statistical analysis of the

results.

By contrast, oligonucleotide hybridisation lends itself to automation and to quantitative analysis [Southern *et al.*, 1992], but it is not well suited to the analysis of variation in the number of repeats in the micro- and minisatellites, as the small fractional change in the number of repeats produces a barely detectable change in signal strength; and of course it would not be possible to distinguish two alleles in the same sample as each would contribute to a single intensity measurement. Thus, many different combinations of alleles would produce the same signal.

Present hybridisation methods are much better suited to analysing variation in the DNA due to point mutation - base substitution deletions and insertions, for which it is possible to design allele specific oligonucleotides (ASOs) that recognise both the wild type and the mutant sequences [Conner *et al.*, 1983]. Thus it is possible in principle, in a relatively simple test, to detect all possible genotypes. However, a problem that arises in practice in the use of oligonucleotide hybridisation is that in some cases the extent of reassociation is barely affected by a mismatched base pair.

#### THE INVENTION

The invention describes a general approach which can be applied to all forms of variation commonly used as DNA markers for genetic analysis. It combines sequence-specific hybridisation to oligonucleotides, which in the preferred embodiment are tethered to a solid support, with enzymatic reactions which enhance the discrimination between matching and non-matching duplexes, and at the same time provide a way of attaching a label to indicate when or which reaction has taken place. Two enzymatic reactions, chain extension by DNA dependent DNA polymerases and DNA strand-joining by DNA ligases, are dependent on perfect matching of sequences at or around the point of extension or joining. As we shall show, there are several ways in which these enzymes

can be used with sequence-specific oligonucleotides to detect variation in target sequences.

In all cases, the sequence to be analysed, the target sequence, will be available as a nucleic acid molecule, and may be a DNA molecule produced, for example, by the polymerase chain reaction. However, the methods are not confined to analysis of DNA produced in this way. In all applications, the target sequence is first captured by hybridisation to oligonucleotides which are preferably tethered to a solid support; for example, the oligonucleotides may be synthesised *in situ* as described [Maskos and Southern, 1992]; or they may be presynthesised and then coupled to the surface [Khrapko *et al*, 1991].

In one aspect of the invention the novelty arises from the exploitation of enzymes in combination with substrates or primers tethered to solid supports. A further novelty exploits the observation that DNA ligases and polymerases can be used to distinguish sequence variants which differ in the number of units of a tandemly repeating sequence. This observation is surprising, as tandemly repeated sequences can form duplex in any register, thus in principle, length variants can form duplexes which match at the ends even when the two strands contain different numbers of repeat units. Although we demonstrate the application of this method in conjunction with tethered oligonucleotides, it should be evident that this reaction could be used to analyse VNTR (variable number tandem repeat) sequences in the liquid phase followed by some other method of analysis, such as gel electrophoresis.

In one aspect the invention provides a method of analysis which comprises: providing a polynucleotide target including a nucleotide at a specified position, and an oligonucleotide probe, tethered to a support, said probe being complementary to the target and terminating at or close to the said specified position; and performing the steps:

- a) incubating the target with the probe to form a duplex,

- b) incubating the duplex under ligation conditions with a labelled oligonucleotide complementary to the target,
- c) and monitoring ligation in b) as an indication of a point mutation at the specified position in the target.

5 In another aspect the invention provides a method of analysis which comprises: providing a polynucleotide target having a variable number tandem repeat section and a flanking section, and an oligonucleotide probe having a section complementary to the repeat section and a flanking section of the target; and performing the steps:

- 10 a) incubating the target with the probe to form a duplex,
- b) incubating the duplex with a labelled oligonucleotide and/or at least one labelled nucleotide under chain extension conditions,
- c) and monitoring chain extension as an indication of the length of the variable number repeat section of the target.

15 A polynucleotide target is provided, in solution when the probe is tethered to a support, and may be DNA or RNA. This polynucleotide target is caused to hybridise with an oligonucleotide probe. The term oligonucleotide is used here, as common terminology for the primers and substrates commonly utilised by polymerase and ligase

20 enzymes. However, the term is used in a broad sense to cover any substance that serves as a substrate for the enzymes, including single stranded chains of short or moderate length composed of the residues of nucleotides or of nucleotide analogues, and also longer chains that would ordinarily be referred to as polynucleotides.

25 The probe may be tethered to a support, preferably by a covalent linkage and preferably through a 5' or 3' terminal nucleotide residue. An array of oligonucleotide probes may be tethered at spaced locations, for example on a derivatised glass surface or the surface of a silicon microchip, or alternatively on individual beads.

30 In another aspect the invention provides an array of

oligonucleotides, for analysing a polynucleotide target containing a variable sequence, in which each component oligonucleotide i) comprises a sequence complementary to the target including an expected variant of the target, and ii) is tethered to a solid support in a chemical orientation which  
5 a) permits duplex formation with the target, and b) permits chain extension only when the sequence of the oligonucleotide matches the variable sequence of the target.

In another aspect the invention provides a set or array of oligonucleotides, for analysing a polynucleotide target containing a variable  
10 number tandem repeat sequence, in which each component oligonucleotide i) comprises a sequence complementary to a part of the target immediately adjacent the repeat sequence, ii) comprises a sequence complementary to the repeat sequence of the target and containing a number of repeats expected in the target, and iii) is configured  
15 in a way that a) permits duplex formation with the target, and b) permits chain extension only when the number of repeats in the oligonucleotide equals or is less than the number of repeats in the target.

In another aspect the invention provides an array of oligonucleotides in which different oligonucleotides occupy different  
20 locations and each oligonucleotide has a 3' nucleotide residue through which it is covalently tethered to a support and a 5' nucleotide residue which is phosphorylated.

The invention also provides a method of making an array of different oligonucleotides tethered to different locations of a support, which  
25 method comprises the steps of: providing a first intermediate oligonucleotide tethered to the support and a second intermediate oligonucleotide in solution, and a third oligonucleotide that is complementary to both the first and second intermediate oligonucleotides, forming a duplex of the third oligonucleotide with the first and second  
30 intermediate oligonucleotides, and ligating the first intermediate

oligonucleotide with the second intermediate oligonucleotide; and repeating the steps with oligonucleotides tethered to different locations of the support

Reference is directed to the accompanying drawings in which  
5 each of Figures 1 to 6 is a series of diagrams illustrating a method according to the invention.

Figure 1 shows detection of point mutation by single base extension.

Figure 2 shows detection of point mutation by hybridisation to  
10 allele specific oligonucleotides and chain extension.

Figure 3A shows detection of point mutation by tag ligation to allele specific oligonucleotides.

Figure 3B shows detection of point mutation by ligation to library of differentially tagged allele specific oligonucleotides.

15 Figure 4A shows analysis of variable numbered tandem repeats by ligation of tag to allelic variants.

Figure 4B shows analysis of variable number tandem repeats by ligation of tag to allelic variants.

Figure 5 shows measurement of variable number tandem  
20 repeats by labelled chain extension.

Figure 6 shows analysis of variable number tandem repeats by ligation of tag followed by chain extension.

~~DETAILED DESCRIPTION~~

25 Detection of point mutation

**1. Single base-specific extension of tethered primers.**

In this application, the tethered oligonucleotide terminates at a position one base before the variable base in the target sequence (Fig.1). A nucleotide precursor triphosphate or dideoxynucleotide  
30 triphosphate, labelled, for example with a fluorescent tag, is added in the

presence of a nucleic acid synthesising enzyme which requires a specific template in order to incorporate the complementary base. In the case of DNA polymerase, the labelled base will be incorporated from a deoxyribonucleotide precursor only if the precursor base is complementary to the base in the target sequence. Thus, mutants will give a negative result.

## II. Chain extension from tethered ASOs.

In this case, the tethered oligonucleotide terminates in a base which is complementary to the variable base in the target sequence. Labelled precursor nucleoside triphosphates and polymerase are added. Polymerisation takes place only if the last base of the primer is complementary to the variable base in the target (Fig.2). Thus, mutants will give a negative result.

## III. Ligation of tag sequences to tethered ASOs.

In this method, the tethered oligonucleotide may end at the variable position in the target sequence, or it may end close to this position. In either case, hybridisation of the target to the tethered ASOs will produce a substrate for ligating a tag oligonucleotide only if the bases at the join are well matched (Fig. 3a). Thus, mutants which are close enough to the joining position to prevent ligation will give a negative result. Alternatively, the tethered oligonucleotide may terminate at the base before the variable position; in this case, the ligation reaction can be carried out using a mixture of tag oligonucleotides, one for each of the possible alternative variants. Each tag would be labelled differently, for example, with a different fluophore, so that those that ligated could be recognised identifying the variant base (Fig. 3b).

### Analysis of VNTR Lengths by Ligation to Anchored VNTRS

In this application "tag" oligonucleotides are ligated to sets of tethered oligonucleotides after hybridisation of the target, which acts as a template to bring the tags and the tethered oligonucleotides together.

5 In Figure 4a, the tethered oligonucleotides comprise three parts, an "anchor" sequence which is common to all members of a VNTR set, which is attached to the solid support and which hybridises to a sequence flanking the variable region, a variable number of the repeated sequence unit, and a distal sequence. Each allele, represented by a  
10 different number of repeats, is located on a different solid support or at a different location on the same solid surface. Hybridisation of the target sequence will produce a series of duplexes, the structures of which depend on the number of units that the target contains. If the number matches the number in a tethered oligonucleotide, the target will meet the end of a tag  
15 when the tag is hybridised to the distal sequence of the tethered oligonucleotide. If the number is greater or smaller, there will be a gap in the duplex which reduces or prevents ligation of the tag.

In Figure 4b, the tethered oligonucleotides comprise two parts: an "anchor" sequence which is common to all members of a VNTR  
20 set and which hybridises to a sequence flanking the repeated region, and a variable number of the repeated sequence unit. Each allele, represented by a different number of repeats, is located on a different solid support, or at a different location on the same solid surface. Hybridisation of the target sequence will produce a series of duplexes, the structures of which depend  
25 on the number of repeat units that the target contains (Fig.4b). If the number matches the number in a tethered oligonucleotide, the latter will meet the end of the tag when the tag is hybridised to its complement in the target sequence, and form a substrate for ligation of the tag. If the number is greater or smaller, there will be a gap in the duplex which reduces or  
30 prevents ligation of the tag.



### Analysis of VNTR Lengths by Chain Extension

The number of repeat units in a VNTR may vary within small limits, and in these circumstances, the method of analysis described above, using a ligase, will be appropriate; in other cases, for example the

5 trinucleotide repeats associated with a number of human inherited disorders, the variation may be too large to analyse in this way. For a number of triplet repeats, the variation can be from around 10-50 in the normal chromosome to more than a thousand in the affected individual (Table 1). It is probably unrealistic to measure such large numbers of

10 repeats using the ligase reaction. In these cases, where the difference between the normal and the mutant allele is large, an alternative is to measure approximately the number of repeat units using labelled precursors with a polymerising enzyme. The enzyme may be either a polymerase, such as DNA dependent DNA polymerase, or a ligase. In the

15 former case, the oligonucleotides have to be tethered at their 5' ends to satisfy the requirement for enzymatic extension by the polymerase. The solid support carries an oligonucleotide anchor complementary to a sequence flanking the repeat unit; for example, the sequence can be that of one of the primers used to amplify the test sequence by the PCR. After

20 hybridisation of the test sequence to the anchor, the repeat insert may be copied by a polymerase or a ligase (Fig. 5) incorporating a labelled precursor. The amount of label incorporated is proportional to the number of repeat units. Incomplete hybridisation of the target to the anchor sequence would give a deceptively low measure of the repeat number.

25 This problem can be overcome by standardising the measurement in one of several possible ways. For example, if the target sequence itself is labelled as shown in Fig. 5, the final measurement will be a ratio of two labels: the target and the incorporated precursor.

Alternatively, in the case of a triplet repeat, incorporation will

30 end at the point in the sequence where the missing precursor base is

needed for further extension; where a ligase had been used to polymerise monomers of the basic repeat unit, this will also end at the end of the VNTR insert. At this point a labelled "capping" sequence can be added by ligation. In such cases, the measurement will be the ratio of cap to  
5 polymer labels.

### Analysis of VNTR Lengths by Combining Ligation and Chain Extension

A more powerful way of analysing VNTRs which may vary in  
10 length over a wide range, would be to test first for ligation to a labelled tag oligonucleotide; this would give the results already described for targets with different lengths of repeats: a negative result where the VNTR lengths were longer or shorter in the target than in the tethered oligonucleotides, and positive results where they were the same. Following ligation, which  
15 we have shown can be made to go to completion, the different length classes will behave differently as substrates for DNA polymerase. Those targets in which the repeat number is less than that of the tethered probe will not act as substrates (Fig. 6c). Targets which have the same number of repeats as the probes will not be elongated by polymerase, because the  
20 ligated tag will block extension (Fig. 6a). The only cases where extension will occur are those for which the targets are longer than the probes (Fig. 6b). If the analysis is done on an array of probes with different numbers of inserts up to a certain limit, there will be a clear indication of the number of repeats in the targets from the ligation results provided they are within the  
25 range of sizes represented in the array of probes. If, within the targets, there is one which is longer than this range, it will show up in the polymerase analysis. This test will be especially useful for the triplet repeats associated with the so-called "dynamic mutations", for example, that which is found in the fragile X mutation, where the size range varies  
30 from ca. 10-1000. It would be difficult to accommodate all of these size

classes on a single array.

## EXPERIMENTAL SUPPORT FOR THE CLAIMS

### Properties of DNA Polymerases and Ligases

5 Most DNA polymerases, reverse transcriptases, some DNA dependent RNA polymerases and ligases can use as substrates one or more oligonucleotides which are bound to a long DNA strand through Watson-Crick base pairing. In the case of polymerases, an oligonucleotide is used as a primer to which the first base in the growing chain is added.

10 In the case of ligases, two oligonucleotides are joined provided that both are paired to the DNA strand and perfectly matched in the base pairs at or close to the junction point. It is these properties that make the enzymes useful for the detection of DNA sequence variation; in particular, the requirement for specific base pairing at the site of extension or joining

15 complements the sequence discrimination that is already provided by the Watson-Crick pairing between the oligonucleotide and the target sequence that is needed to form a stable duplex. Thus, it has been found that discrimination by hybridisation alone is most sensitive if the variant base(s) is (are) close to the middle of the oligonucleotide. By contrast, for the

20 enzymes, discrimination is highest if the variant mismatching bases are close to the end where the extension or join takes place. Together, hybridisation under stringent conditions and enzymatic extension or joining provide greater discrimination than either alone, and several methods have been developed to exploit this combination in systems for genetic analysis

25 [references in cotton, 1993]. The hybridisation and the enzyme reaction are normally carried out in solution, following which the product is captured on a solid support, or separated by gel electrophoresis for detection and/or measurement.

In one embodiment, the invention described here employs

30 oligonucleotides coupled to a solid surface, so that the advantages of

working in mixed phase are brought to all steps: hybridisation, enzymatic extension or joining, and detection. This provides great sensitivity and convenience. As many different oligonucleotides can be bound to one surface in an array, it enables many different sequences to be analysed together, in a single reaction; this also ensures that all reactions are carried out under identical conditions, making comparisons more reliable.

### Support-bound Oligonucleotides

Two different methods have been developed for making oligonucleotides bound to a solid support: they can be synthesised *in situ*, or presynthesised and attached to the support. In either case, it is possible to use the support-bound oligonucleotides in a hybridisation reaction with oligonucleotides in the liquid phase to form duplexes; the excess of oligonucleotide in solution can then be washed away. Hybridisation can be carried out under stringent conditions, so that only well-matched duplexes are stable. When enzymes are to be used, the chemical orientation of the oligonucleotide is important; polymerases add bases to the 3' end of the chain; ligases join oligonucleotides which are phosphorylated at the 5' end to those with a 3'-OH group. Oligonucleotides tethered to the solid substrate through either end can be made *in situ* by using the appropriate phosphoramidite precursors [references in Beaucage and Iyer, 1992]; or presynthesised oligonucleotides can be fixed through appropriate groups at either end. We will demonstrate that oligonucleotides can be phosphorylated at the 5' end *in situ* using ATP and polynucleotide kinase, or they may be phosphorylated chemically [Horn and Urdea, 1986].

### Tethered Oligonucleotides as Substrates for DNA Modifying Enzymes

The applications envisaged here require that the oligonucleotides tethered to the solid substrate can take part in reactions catalysed by DNA polymerases and ligases.

### DNA polymerase

~~The M13 sequencing primer -~~

5'-GTAAAACGACGGCCAGT-3' - attached to aminated polypropylene through its 5' end was synthesised as described. A solution of M13 DNA (single-strand, replicative form, 0.1  $\mu$ l, 200 ng/ $\mu$ l) was applied in two small spots to the surface of the derivatised polypropylene. A solution containing three non-radioactive deoxyribonucleotide triphosphates, dATP, dGTP, TTP (10  $\mu$ mol each),  $\alpha^{32}$ P-dCTP (10  $\mu$ Ci), Taq DNA polymerase and appropriate salts, was applied over a large area of the polypropylene, including the area where the M13 DNA had been spotted. The polypropylene was incubated at 37°C for 1 hr in a vapour saturated chamber. It was then washed in 1% SDS at 100°C for one minute, and exposed to a storage phosphor screen for one minute and scanned in a phosphorimager. The regions where the DNA had been applied showed a high level of radioactivity, against a low background where no DNA had been applied. This experiment shows that oligonucleotides tethered to a solid support can act as primers for DNA-dependent synthesis by DNA polymerase, as required for applications using this enzyme for mutation detection.

Experiments described below show that both polynucleotide kinase and DNA ligase can be used to modify oligonucleotides tethered to a solid support. There are several ways in which phosphorylated oligonucleotides and the ligase reaction can be used to detect sequence variation.

### Methods for Making Arrays of Sequence Variants.

#### 1. Allele specific oligonucleotides for point mutations.

For the preferred embodiment, it will be necessary to use oligonucleotides tethered to a solid support. The support may take the form of particles, for example, glass spheres, or magnetic beads. In this

case the reactions could be carried out in tubes, or in the wells of a microtitre plate. Methods for both synthesising oligonucleotides and for attaching presynthesised oligonucleotides to these materials are known [Stahl *et al.*, 1988]. Methods for making arrays of ASO's representing point mutations were described in patent application PCT/GB89/00460 and in Maskos and Southern (1993). We also demonstrated how oligonucleotides tethered to a solid support in an array could distinguish mutant from wild type alleles by molecular hybridisation.

For the present invention, the same methods could be used to create oligonucleotide arrays of ASOs, but in order that they can be used as substrates for the enzymes, they need to be modified; for ligation, it may be necessary to phosphorylate the 5' end; for extension by polymerase it will be necessary to attach to oligonucleotides to the solid substrate by their 5' ends.

## 2. Arrays for scanning regions for mutations.

It is often desirable to scan a relatively short region of a gene or genome for point mutations: for example, many different sites are mutated in the CFTR gene to give rise to cystic fibrosis; similarly, the p53 tumour suppressor gene can be mutated at many sites. The large numbers of oligonucleotides needed to examine all potential sites in the sequence can be made by efficient combinatorial methods [Southern *et al.*, 1994]. A modification of the protocol could allow such arrays to be used in conjunction with enzymes to look for mutations at all sites in the target sequence.

## 3. VNTRs.

The most commonly used VNTRs are repeats of very short units, typically mono to tetranucleotides. However, there is another class, the minisatellites, in which the repeat unit is somewhat longer, up to 20 or

more nucleotides. The short repeats may be made using chemical synthesis; in the case of inserts with large numbers of repeat units, it would be more economical to use a synthetic route which used the repeat unit as a reactant, rather than building them up one base at a time; such methods have been used to make polynucleotides by chemical synthesis. An attractive alternative would be to build the repeat units by ligating the monomer units; they could be added stepwise, one unit at a time provided a method could be found to block one end to prevent polymerisation; for example the oligonucleotide building block may be terminated by a hydroxyl group, which is then phosphorylated after ligation so that the unit becomes an acceptor for the next one; the monomer may have a phosphate group protected by a cleavable group, such as a photocleavable group, which can be removed after ligation to allow a subsequent ligation [Pillai, 1980]. A second alternative, which would be especially favourable for longer units such the minisatellites, would be to attach either cloned or enzymatically amplified molecules to the solid support. For example, each variant sequence could be amplified by the PCR, using a biotinylated oligonucleotide for one of the primers. The strand starting with this group could then be attached to a streptavidin coated surface, and the other strand removed by melting [Stahl *et al.*, 1988].

### EXAMPLE 1

#### 25 DEMONSTRATION OF THE ANALYSIS OF LENGTH POLYMORPHISM BY LIGATION TO AN ARRAY OF VNTRs

30 An array of VNTRs was made as described in Fig. 4b, in which the anchor sequence was 5'-tctagtgggtgatcaaggc-3'. The repeat unit was 5'-cttt-3', stripes, ca 3 mm wide, of sequence variants of the form: Anchor-Repeat<sub>N</sub>, with N = 4-10, were made as stripes on the surface of a

Cont  
ag

sheet of polypropylene. The synthesis was carried out using 3'-deoxyribophosphoramidites, this chemical orientation produces oligonucleotides tethered through their 3' ends to the polypropylene, and a free 5' hydroxyl group. To create a substrate for ligation, this OH group

5 was phosphorylated by immersing a strip of the polypropylene (3mm x 18mm), carrying the array of oligonucleotides, in 0.5 ml of a solution containing 4mM ATP and 77.6 units of polynucleotide kinase with buffer and Mg<sup>++</sup> according to the supplier's instructions. The reaction was left for 6 hours at 37°C; the strip was removed and immersed in boiling water to

10 kill the polynucleotide kinase. The target sequences, which are complementary to elements of the array oligonucleotides and to the ligation tag, 5'-Anchor-Repeat<sub>10</sub>-Tag and 5'-Anchor-Repeat<sub>5</sub>-Tag were added to 0.5 ml of a solution, preheated to 95°C, containing the tag, 5'-gtggtcactaaagtcttctgct-3', which had been labelled at its 5' end using

15 polynucleotide kinase and <sup>33</sup>P-gamma-ATP, thermal ligase (500 units), and buffer and salts according to the supplier's instructions. The polypropylene strip was immersed in the hot solution, which was then allowed to cool to 68°C, and left at this temperature for 16 hours. The polypropylene strip was removed and placed in 25% formamide at 95°C for 5 minutes, rinsed

20 in water at the same temperature, dried and exposed to a storage phosphor screen, from which an image of the radioactivity was collected. The results showed counts close to background over most of the array; counts on Anchor-Repeat<sub>5</sub> and Anchor-Repeat<sub>10</sub> were more than five times those over adjacent cells in the array. This experiment indicates that the

25 ligase is able to distinguish length variants of the repeat sequence and gives optimum ligation only when the number of repeats in the target matches that in the allele specific oligonucleotide in the array. Thus, it should easily be possible to detect the two allelic variants in a heterozygote.

30 Ligation and/or polymerisation is possible if the



oligonucleotide is tethered through the 5' end. The oligonucleotides can be synthesised *in situ* using deoxyribophosphoramidites with 3' dimethoxytrityl groups and 5' phosphoramidite (reverse phosphoramidites). It is unnecessary to phosphorylate the tethered polynucleotide for ligation assays since the phosphate needed for ligation is provided by the tag oligonucleotide.

## EXAMPLE 2

### Analysis of VNTR lengths

10 An array of VNTR's was made as described in Figure 4B, with the oligonucleotides anchored through their 5' ends. The repeat unit was 5' tcca and the anchoring sequence 5' ctattccctca. Stripes 6mm wide of sequence variants of the form: Anchor-Repeat<sub>N</sub> where N = 4-8 were made on the surface of a sheet of polypropylene using "reverse" 15 phosphoramidite monomers.

### Analysis by ligation

20 A strip of the array (30mm x 2mm) was immersed in a solution of 600pmols of the target oligonucleotide 5' cacagactccatgg(tgaa)<sub>6</sub>tgagggasataag, 1.4 pmol of oligo 5' ccatggagctctgtg (labelled at its 5' end using polynucleotide kinase and <sup>33</sup>P gamma ATP) and buffer and salts according to the suppliers instructions, the total volume being 293 µl. The solution was heated to 65°C and 7 µl of Tth DNA ligase added. The reaction was then cooled to 37°C and left at that 25 temperature for 18hrs. After removal from the reaction solution the strip was washed in T.E. buffer, blotted dry and exposed to a storage phosphor screen from which an image of the radioactivity was taken. The results showed that the target sequence ligated to the correct sequence with a higher yield than to the shorter and longer sequences in adjacent cells of 30 the array.

### EXAMPLE 3

#### Analysis by ligation and polymerisation

A strip of the array (30mm x 2mm) from Example 2 was added to a solution of 200 pmols of the target oligonucleotide 5'

- 5 cacagactccatgg(tgaa)<sub>6</sub>tgagggaaataag, 200 pmol of oligo 5' ccatggagtctgtg (chemically phosphorylated at the 5' end) with buffer and salts according to the suppliers instructions, the total volume being 243 µl. The solution was heated to 85°C and cooled to 37°C over a period of 30 mins. 7 µl of Tth DNA ligase was added and the reaction mixture heated at 34°C for 17hrs.
- 10 The strip was removed and added to a solution of 8 mM DTT, 3.3 pmol <sup>32</sup>P alpha dTTP, 13 units sequenase version 2.0 and buffer and salts according to the suppliers instructions. The total volume was 250 µl. After heating at 37°C for 3 hrs the strip was removed from the reaction solution, washed in TE buffer, blotted dry and exposed to a storage phosphor
- 15 screen from which an image of the radioactivity was taken. The results showed counts equal to background over the area of the array where repeats were equal in length or greater than the repeat length of the target, with 20 times the signal in the areas where the repeat length of the array was shorter than the repeat length in the target.

20

### EXAMPLE 4

#### Analysis by polymerisation

- Two types of polymerase analysis were carried out where reporter nucleotides were chosen, in one case to identify the correct repeat
- 25 length, and in the other case to identify shorter repeat lengths. This is made possible when the repeat sequence comprises less than all four bases.

- In the former case a base is chosen which is present in the repeat sequence and is different from the first base in the flanking
- 30 sequence. In the latter case a base is chosen to be complementary to the

first base in the flanking sequence which is absent from the repeat.

A strip of the array from Example 2 (30mm x 2mm) was added to a solution of 500 pmols of the target oligonucleotide 5' cacagactccatgg(tgaa)<sub>6</sub>tgaggggaaataag in buffer and salts at a concentration 1.09 times the suppliers instructions, the total volume being 275 µl. The solution was heated to 75°C for 5 minutes and cooled to 37°C over a period of 25mins. The solution was removed and added to 3.3 pmols <sup>32</sup>P alpha dCTP, 5 µl 1M DTT. 13 units of Sequenase version 2.0 and water to give a final volume of 295 µl. This solution was added to the array and heated at 37°C for 15hrs 40mins. The polypropylene strip was removed, washed in water and exposed to a storage phosphor screen. The results showed counts of 5 times more for the correct sequence than shorter sequences and twice for the correct sequence compared with longer repeats.

A similar strip of the array (30mm x 2mm) was added to a solution of 500 pmols of the target oligonucleotide 5' cacagactccatgg(tgaa)<sub>6</sub>tgaggggaaataag in buffer and salts 1.09 times the suppliers instructions, the total volume being 275 µl. The solution was heated to 75°C for 5 mins and cooled to 37°C over a period of 25 mins. The solution was removed and added to 3.3 pmols <sup>32</sup>P alpha dTTP, 5 µl 1M DTT, 13 units of Sequenase version 2.0 and water to give a final volume of 295 µl. This solution was added to the array and heated at 37°C for 15hrs 40mins. The polypropylene strip was removed, washed in water and exposed to a storage phosphor screen. The results showed counts of 4.5 times more for the shorter array sequences than the correct and longer repeat lengths.

#### EXAMPLE 5

##### VNTR analysis by ligation

In an experiment similar to the one described in Example 1, an array was created using the human fes/fps locus sequence as a target.

Get all

The anchor sequence 5' agagatgtagtctcattcttcgccaggetgg 3' was the actual flanking sequence to the att repeats of the fes/fps microsatellite (EMBL Accession No X06292-M14209 M14589) as it occurs in human genomic DNA. Using a target oligonucleotide representing the 10 repeat allele and ligating a <sup>32</sup>P labelled 5' flanking sequence (5' g gag aca agg ata gca gtt c 3') and doing a similar experiment to that described above, the resulting radioactivity on the anchor-repeat<sub>10</sub> cell was over 10 fold that on adjacent cells in the array.

10

### EXAMPLE 6

#### Demonstration of stepwise ligation of oligonucleotides bound to a solid support.

from 12

A primer oligodeoxynucleotide - 5' PO<sub>4</sub> gta aaa cga cgg cca gt 3', attached to aminated polypropylene through its 3' end, was synthesised and phosphorylated as described. A small square (2mm x 2mm) piece of this material was placed in standard ligation buffer, with template oligonucleotide 5'tcg ttt tac cgt cat gcg tcc tct ctc 3' (250 nM) and a protected ligator oligonucleotide 5' NB PO<sub>4</sub> cgc atg acg 3' (250 nM) and <sup>32</sup>P labelled extender oligonucleotide 5' gag aga gga 3', where NB is a protecting group based on a photocleavable o-nitrobenzyl derivative. The NB protected phosphate of the ligator oligonucleotide had previously been shown to be unable to take part in the ligation reaction. The NB group had also been shown to be removable by uv light to leave a fully functional phosphate group. To this mixture was added thermus thermophilus DNA ligase (Advanced Biotechnologies) 25u and the reaction incubated at room temperature for 6 hours. The mixture was then irradiated with uv light (20 minutes room temperature) and incubated for a further 12 hours. The polypropylene patch was then washed with 30% formamide at 95°C for 5 minutes, and exposed to a storage phosphor screen for 24 hours and scanned in a phosphorimager. The patch showed

GA  
Q12  
a level of radioactivity 50 fold higher than a patch treated in a similar fashion but without addition of the central "ligator" oligonucleotide. In a similar experiment using a phosphorylated ligator oligonucleotide a similar amount of radioactive extender oligonucleotide became covalently  
5 attached to a third polypropylene/oligonucleotide primer square.

### EXAMPLE 7

#### Demonstration of point mutation analysis by ligation to allele specific oligonucleotides attached to solid supports

10 Four tethered ASOs 5' (gca or t)ag aga gga 3', differing only at their 5' base, were synthesised as described above, with the 3' end attached to aminated polypropylene. Phosphorylation was carried out as described and four squares of polypropylene carrying each ASO were  
15 placed in standard ligation buffer along with complementary target oligonucleotide 5' tcc tct ctc cgt cat gcg tat cgt tca at 3' (250 nM). After addition of  $^{33}\text{P}$  labelled ligator oligonucleotide 5' cgc atg acg 3' (10 nM) and thermus thermophilus DNA ligase (100u), the mixture was incubated at  
20 37°C for 18 hours. The ASO which was fully complementary to the target oligonucleotide was found to have acquired 100-fold greater radioactivity through ligation of the labelled ligator than the non-complementary ASOs.

### EXAMPLE 8

#### Demonstration of DNA ligation specificity

25 In a model experiment to assess the specificity of TTh DNA ligase, ligator and extender deoxyoligonucleotides were ligated together by means of hybridisation to an oligonucleotide template and ligation by DNA ligase.

30 Template oligonucleotide 5' tcc tct ctc cgt cat gcg tat cgt tca at 3' (250 nM), phosphorylated,  $^{33}\text{P}$  labelled, extender oligonucleotide

GA  
Q14

5' PO<sub>4</sub>gag-aga gga 3' (10 nM) and ligator sequence 5' gca gta cg 3' (250 nM) were mixed together in standard ligation buffer with DNA ligase 25 u. This mixture was incubated at 35°C. Samples of this mixture were removed and the reaction stopped by addition of formamide at 15, 30, 60, 120, and 240 minutes. The ligated and unligated products were separated by 20% denaturing polyacrylamide gel electrophoresis. The gel was exposed to a phosphor screen for 18 hours and scanned by a phosphorimager. The relative proportions of ligated to unligated products of the reaction were then measured. 50% of the extender sequence had been ligated to the ligator sequence in 30 minutes. By comparison in a similar experiment with ligator 5' gca tga ag 3' after 30 minutes only 1% of the extender sequence had become ligated.

Other polymerases and ligases such as Taq polymerase, Thermosequenase, T4 DNA ligase and *E. Coli* DNA ligase have also been shown to be useful in experiments similar to those described above.

### REFERENCES

1. Beaucage, S. L. and Iyer, R. P. (1992). Advances in the synthesis of oligonucleotides by the phosphoramidite approach. Tetrahedron 48: 2223-2311.
2. Conner, B. J., Reyes, A. A., Morin, C., Itakura, K., Teplitz, R. L., and Wallace, R. B. (1983). Detection of sickle cell  $\beta^s$  globin allele by hybridization with synthetic oligonucleotides. Proc. Natl. Acad. Sci. USA 80: 278-282.
3. Cooper, D. N., and Krawczak, M. (1989). The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. Hum. Genet. 85: 55-74.
4. Cotton, RG, (1993) Current methods of mutation detection. Mutation Research 285: 125-144.

5. Horn, T., and Urdea, M. (1986) Chemical phosphorylation of oligonucleotides. *Tetrahedron Letters* **27**: 4705.
6. Khrapko, K. R., Lysov, Yu. P., Khorlyn, A. A., Shick, V. V., Florentiev, V. L., and Mirzabekov. (1989). An oligonucleotide hybridization  
5 approach to DNA sequencing. *FEBS Lett.* **256**: 118-122.
7. Krawczak M. and Schmidtke, J. (1994). DNA fingerprinting. BIOS Scientific Publishers.
8. Maskos, U., and Southern, E.M., (1993) A novel method for the analysis of multiple sequence variants by hybridisation to  
10 oligonucleotides. *Nucleic Acids Research*, **21**: 2267-2268.
9. Pillai, V. N. R.(1980). Photoremovable protecting groups in organic chemistry *Synthesis* **39**: 1-26.
10. Southern, E. M. (1988). Analyzing Polynucleotide Sequences. International Patent Application PCT/GB89/00460.
- 15 11. Southern, E.M., Maskos, U. and Elder, J.K. (1992). Analysis of Nucleic Acid Sequences by Hybridization to Arrays of Oligonucleotides: Evaluation using Experimental Models. *Genomics* **12**: 1008-1017.
12. Southern, E. M., Case-Green, Elder, J.K. Johnson, M., Mir, K.U., Wang, L., and Williams, J.C. (1994). Arrays of complementary  
20 oligonucleotides for analysing the hybridisation behaviour of nucleic acids. *Nucleic Acids Res.* **22**., 1368-1373.
13. Stahl, S., Hultman, T., Olsson, A., Moks, T.D and, Uhlen, M. (1988) Solid phase DNA sequencing using the biotin-avidin system. *Nucleic Acids Res.* **16**: 3025-38.
- 25 14. Veerle, A.M.C.S., Moerkerk, P.T.M.M., Murtagh, J.J., Jr., Thunnissen, F.B.J.M (1994) A rapid reliable method for detection of known point mutations: Point-EXACCT. *Nucleic Acids Research* **22**: 4840-4841.





## DNA sequences of triplet repeats

Condition	Repeat	Normal	Preexpansion	Expanded	Reference
FRAXA	CGG/CCG	10-50	38-50	200-1000	Verkerk et al. 1991
FRAXE	CGG/CCG			200-1000	Knight et al. 1993
FRAXF	GCC/CGG	6-18	?	300-500	Parrish et al. 1994
FRAX16A	CGG/CCG			1000-2000	Nancarrow et al. 1994
SBMA	CAG	11-31	?	40-62	Tilley et al. 1994
Huntington	CAG	11-34	30-34	42-100	Huntington group 1993
SCA1	CAG	25-36	?	43-81	Orr et al. 1993
DRPLA/HRS	CAG			≤100	Burke et al. 1994
Machado-Joseph	CAG/CTG	~26	?	68-79	Kawaguchi et al. 1994

PCT/GB96/00848